Emilie Doyle

ECE 3522: Stochastic Process in Signals and Systems

Department of Electrical and Computer Engineering, Temple University, Philadelphia, PA 1912

# Problem Statement

The overall objective of this assignment is to gain more knowledge and practice with linear regressions and histograms, especially that creation of these plots through the use of MATLAB. The given information was the exact same data from the first computer assignment- Google stock price data since 2004 in the form on a Microsoft Excel table, and a \*.raw audio signal. The objective for the first part of this assignment was build off of the windows and frames analysis of the first assignment, specifically the data acquired from using a frame size of 1 and a window size of 7 to find the mean. The concept of linear regression had to be researched and implemented on the Google stock data. This regression model was then compared to the mean that resulted from the windows and frames analysis. The second part of the assignment focused on the speech signal. The amplitude of the audio signal was found and plotted as a histogram with bin size of 10 and spanning from -32767 to +32767. The samples also had to be normalized in order for them to be accurately compared and outliers to be withheld. Lastly, the cumulative distribution had to be plotted.

# Approach and Results

For part 1, I reused code from the first computer assignment in order to get the data in from the Excel table into an array. I also used the windows and frames code from the first computer assignment to find the mean data pertaining to a frame of 1 and a window of 7. I isolated the necessary window and frame values from the array of values used in assignment 1 so the loop would only process the values needed for this assignment. In the same figure, I used subplot to plot the window and frame analysis mean values, followed by the actual closing price, and then a plot that contained the mean value plotted on the same graph as a copy of the actual values. I maintained line colors across the graphs for visual clarity.

. 

Figure : Windows and Frames Mean, Actual Data, Both









Figure : Zoomed In Views of Mean Price and Signal

After these values were found the regression model had to be found too. In order to find the regression model, I used the ‘plotregression’ function in MATLAB. This function allows for the linear regression to be plotted with the actual data. It is important to note that the R value for the linear regression model is 0.86604.



Figure : Linear Regression Model of Mean



Figure : Zoomed in Linear Regression Model

For part 2, I reused code from the first computer assignment in order to get the audio signal into MATLAB as an array. The first objective of part 2 was to create a histogram of the amplitude of the signal values. I used the histogram function to create a histogram of the signal’s array of data. I established the specifics of the edges of the plot and the size of the bins through creating an array that stretched from the given negative boundary to the given positive boundary and incremented by the correct bin size.



Figure : Histogram of Audio Signal Content



Figure : Close Up of Histogram

The next objective was to plot the cumulative distribution function of the audio signal content. I was able to accomplish this quite easily through the use of the MATLAB ‘cdfplot’ function, which took in the signal content array and created the cumulative distribution function plot.



Figure : Cumulative Distribution Function Plot of Audio Signal

# MATLAB Code







# Conclusions

The objective of this assignment was to learn more about histograms and regression models, and the implementation of these concepts in MATLAB. The regression model of the Google stock prices was quite accurate, and quite similar to the mean plot that was found through the windows and frames analysis. It is important to note that the R value for the linear regression model is 0.86604. The value of R determines the correlation between the data, essentially how accurate the regression model is. The closer the R value is to 1, the more accurate it is to the data. The fact that it is positive indicates that it is a positive relationship, which is if one value increases the other does too. For the second part of the assignment, the values produced by the histogram seem consistent with the data found from the first computer assignment- the bounds of the range appear to match those found. The majority of the content spans the center range of the plot, which implies consistency over a range of values. Additionally the cumulative distribution function makes sense because normal CDFs are the summation of the probability of each value, so over the span of one data set the CDF should rise to 1, which this plot does.